

高等学校计算机课程规划教材



数据库 应用技术

喻梅 主编 / 王庆节 副主编 / 于健 王赞 编

 天津大学出版社
TIANJIN UNIVERSITY PRESS

014002969

高等学校计算机课程规划教材

TP311.138SQ-43
101

要容内

数据库应用技术

喻梅 主编

王庆节 副主编

于健 王赞 编



TP311.138SQ-43
101



天津大学出版社

TIANJIN UNIVERSITY PRESS



北航

C1688619

内容提要

全书以 SQL Server 2005 数据库管理系统为基础,介绍数据库系统的基本原理及应用技术。

全书首先介绍了数据库系统的基本概念和关系数据库的理论,接着简单介绍了 SQL Server 2005 的性能、安装、配置和使用等基本知识以及所提供的各种常用管理工具,然后讲述了创建和管理数据库的各种操作以及在数据库中创建和管理表的各种操作。讨论了使用选择查询从数据库中检索数据和更新数据库中的数据,使用索引来提高检索效率、管理和使用视图的各种操作,创建及维护数据完整性,使用 Transact-SQL 语言进行程序设计。使用存储过程和触发器来控制数据库中的数据。介绍了 SQL Server 2005 的安全性管理和权限设置以及备份和还原数据库的方法。

本书可以作为高等院校相关专业数据库应用技术课程的教材,同时也可供从事数据库研究和使用的 SQL Server 2005 进行数据库系统开发的计算机专业人员参考。

图书在版编目(CIP)数据

数据库应用技术/喻梅主编. —天津:天津大学出版社,
2013. 6

高等学校计算机课程规划教材

ISBN 978-7-5618-4745-9

I. ①数… II. ①喻… III. ①关系数据库系统—高等
学校—教材 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2013)第 182599 号

出版发行 天津大学出版社
出 版 人 杨欢
地 址 天津市卫津路 92 号天津大学内(邮编:300072)
电 话 发行部:022-27403647
网 址 publish.tju.edu.cn
印 刷 河间市新诚印刷有限公司
经 销 全国各地新华书店
开 本 185mm×260mm
印 张 23
字 数 574 千
版 次 2013 年 6 月第 1 版
印 次 2013 年 6 月第 1 次
定 价 42.00 元

凡购本书,如有缺页、倒页、脱页等质量问题,烦请向我社发行部门联系调换

版权所有 侵权必究

前 言

由于计算机技术的进步,使得关系型数据库管理系统软件得到了快速发展,尤其是微软公司的 SQL Server 数据库系统,从 SQL Server 7.0、SQL Server 2000 发展到现在的 SQL Server 2005,并且随着版本的不断升级,功能也越来越强大。可视化数据库设计与管理是当前的技术热点,本书是针对目前流行的 SQL Server 2005 而编写的数据库应用基础教材。

SQL Server 是一个全面的、集成的、端到端的数据解决方案,它为组织中的用户提供了一个更安全可靠和更高效的平台,用于企业数据和商务智能(BI)应用。SQL Server 2005 为 IT 专家和信息工作者带来了强大的、熟悉的工具,同时降低了在从移动设备到企业数据系统的多平台上创建、部署、管理和使用企业数据和分析应用程序的复杂性。通过全面的功能集与现有系统的互操作性以及对日常任务的自动化管理能力,SQL Server 2005 为不同规模的企业提供了一个完整的数据解决方案。

SQL Server 2005 包含的内容很多,编写本书时,在具体内容的安排上,从应用的角度出发,以实用性为重点,侧重于操作和应用所必需的基础知识,使读者着重理解构建、管理数据库、对数据操作的思路和方法以及利用 SQL Server 2005 所提供的工具有效地管理数据。在取材的深度和广度方面进行了精心的优化筛选,注意内容简练,精心设计实例,语言通俗易懂。为了便于学习掌握,在例题的安排上注意了连续性,各章相对独立,又互为补充,使学生可以在有限的学时内更好地掌握数据库的实用技术。

本书取材注重实用,内容由浅入深,语言简练,逻辑性强,适用于教学。本书可作为高等院校大学本科相关专业数据库应用技术课程教材的使用,同时也可供从事数据库研究和使用的 SQL Server 2005 进行数据库系统开发的计算机专业人员参考。

本书第1~3,6,7章由喻梅、王赞编写,第4~5,8~10章由王庆节编写,第11~13章由于健编写。书中实例的上机验证及截图由胡悦、雷霆、张旭、吕方完成。在本书的撰写过程中,得到了葛卫民老师的指导,在此表示衷心的感谢。

由于时间仓促,作者水平有限,书中难免有不当之处,敬请读者批评指正。

编者

2013年4月

目 录

第1章 数据库系统概论	(1)
1.1 信息、数据与数据处理	(1)
1.1.1 信息与数据	(1)
1.1.2 数据处理	(1)
1.2 数据管理技术的发展	(2)
1.2.1 人工管理阶段	(2)
1.2.2 文件系统阶段	(2)
1.2.3 数据库系统阶段	(4)
1.3 数据模型	(5)
1.3.1 数据描述的三个领域	(5)
1.3.2 数据模型	(7)
1.3.3 概念数据模型	(8)
1.3.4 结构数据模型	(11)
1.4 数据库的体系结构	(14)
1.4.1 数据库系统的模式结构	(14)
1.4.2 三级模式结构	(15)
1.4.3 两级模式映像及数据独立性	(16)
1.5 数据库系统	(17)
1.5.1 数据库系统的组成	(17)
1.5.2 数据库管理系统	(18)
1.5.3 数据库管理员	(19)
习 题	(20)
习题答案	(22)
第2章 关系数据库系统	(23)
2.1 关系模型的基本概念	(23)
2.1.1 关系模型的基本术语	(23)
2.1.2 关系的定义和性质	(25)
2.1.3 关系模型的三要素	(25)
2.2 关系代数	(26)
2.2.1 传统的集合运算	(27)
2.2.2 专门的关系运算	(28)
2.2.3 关系代数表达式及其应用实例	(31)
2.3 关系规范化	(32)
2.3.1 关系模式的设计问题	(32)

2.3.2 函数依赖	(34)
2.3.3 关系模式的范式与规范化	(34)
习 题	(37)
习题答案	(39)
第3章 数据库设计	(40)
3.1 数据库设计概述	(40)
3.1.1 数据库设计的内容	(40)
3.1.2 数据库设计的方法	(41)
3.1.3 数据库设计的步骤	(42)
3.2 需求分析	(43)
3.2.1 需求分析的任务	(43)
3.2.2 需求分析的基本步骤	(44)
3.3 概念设计	(44)
3.3.1 概念设计的目标和策略	(44)
3.3.2 采用 E-R 方法的数据库概念设计	(45)
3.4 逻辑设计	(47)
3.4.1 逻辑设计的步骤	(47)
3.4.2 E-R 模型向关系数据模型的转换	(48)
3.4.3 关系数据库的逻辑设计	(49)
3.5 物理设计	(51)
3.5.1 物理设计的内容	(51)
3.5.2 物理设计的性能	(52)
3.6 实现与维护	(53)
3.6.1 数据库的实现	(53)
3.6.2 数据库的其他设计	(54)
3.6.3 数据库的运行与维护	(54)
习 题	(55)
习题答案	(56)
第4章 SQL Server 2005 概述	(57)
4.1 SQL Server 2005 简介	(57)
4.1.1 SQL Server 的发展	(57)
4.1.2 SQL Server 的特点	(57)
4.1.3 SQL Server 2005 的新增功能	(58)
4.1.4 SQL Server 2005 的主要技术	(59)
4.2 SQL Server 2005 的安装准备	(60)
4.2.1 SQL Server 2005 的版本	(60)
4.2.2 SQL Server 2005 的安装环境	(61)
4.3 SQL Server 2005 的安装	(63)
4.3.1 安装 SQL Server 2005 组件	(63)

4.3.2 SQL Server 2005 实用工具	(75)
4.4 SQL Server Management Studio 的使用	(81)
4.4.1 SQL Server 2005 服务器的管理	(81)
4.4.2 SQL Server 2005 的管理平台	(81)
4.5 SQL 语言概述	(86)
4.5.1 SQL 语言的发展	(86)
4.5.2 SQL 语言的特点	(86)
4.5.3 SQL 语言的功能	(87)
4.5.4 Transact-SQL	(87)
习 题	(87)
习题答案	(89)
第5章 SQL Server 数据库的创建与管理	(90)
5.1 SQL Server 数据库概述	(90)
5.1.1 数据库引擎	(90)
5.1.2 文件和文件组	(90)
5.1.3 事务日志	(93)
5.1.4 数据库快照	(94)
5.2 系统数据库	(95)
5.3 创建数据库	(98)
5.3.1 使用图形工具创建数据库	(98)
5.3.2 用 Transact-SQL 命令创建数据库	(100)
5.4 管理数据库	(106)
5.4.1 查看数据库信息	(106)
5.4.2 打开数据库	(107)
5.4.3 修改数据库	(108)
5.4.4 删除数据库	(109)
习 题	(110)
习题答案	(111)
第6章 数据表的创建与管理	(112)
6.1 数据表的建立	(112)
6.1.1 数据类型	(112)
6.1.2 数据表的创建	(116)
6.1.3 特殊类型表	(123)
6.2 数据表的修改	(124)
6.2.1 查看数据表	(124)
6.2.2 修改数据表	(125)
6.2.3 删除数据表	(130)
习 题	(131)
习题答案	(132)

第7章 数据查询与更新	(133)
7.1 数据查询	(133)
7.1.1 Transact-SQL 查询语句	(133)
7.1.2 SELECT 子句	(134)
7.1.3 FROM 子句	(143)
7.1.4 WHERE 子句和 HAVING 子句	(145)
7.1.5 GROUP BY 子句	(156)
7.1.6 ORDER BY 子句	(158)
7.1.7 联接查询	(160)
7.1.8 子查询	(167)
7.2 数据更新	(178)
7.2.1 插入数据	(179)
7.2.2 更新数据	(184)
7.2.3 删除数据	(187)
习题	(191)
习题答案	(192)
第8章 索引与视图	(193)
8.1 使用索引	(193)
8.1.1 索引类型	(193)
8.1.2 索引设计准则	(194)
8.1.3 创建索引	(195)
8.1.4 修改索引	(199)
8.1.5 删除索引	(205)
8.2 使用视图	(206)
8.2.1 视图的作用	(207)
8.2.2 创建视图	(209)
8.2.3 修改视图	(212)
8.2.4 删除视图	(218)
习 题	(219)
习题答案	(221)
第9章 数据完整性	(222)
9.1 数据完整性	(222)
9.1.1 关系数据完整性	(222)
9.1.2 SQL Server 2005 中的数据完整性	(222)
9.2 约束	(223)
9.2.1 主键约束	(224)
9.2.2 外键约束	(225)
9.2.3 UNIQUE(唯一)约束	(227)
9.2.4 检查约束	(228)

9.2.5 默认约束	(229)
9.3 规则	(230)
9.3.1 创建规则	(231)
9.3.2 查看规则	(231)
9.3.3 绑定与解除规则	(232)
9.3.4 删除规则	(232)
9.4 默认值	(233)
9.4.1 创建默认值	(233)
9.4.2 绑定与解除默认值	(234)
9.4.3 删除默认值	(234)
习 题	(235)
习题答案	(236)
第10章 Transact-SQL 程序设计	(237)
10.1 Transact-SQL 语言基础	(237)
10.1.1 Transact-SQL 语言的编程功能	(237)
10.1.2 标识符	(237)
10.1.3 注释	(238)
10.1.4 语句块	(238)
10.2 表达式	(239)
10.2.1 常量	(239)
10.2.2 变量	(240)
10.2.3 运算符	(243)
10.3 函数	(244)
10.3.1 内置函数	(244)
10.3.2 用户定义函数	(248)
10.4 流程控制语句	(253)
10.4.1 批处理	(254)
10.4.2 选择语句	(255)
10.4.3 循环语句	(259)
10.5 游标	(260)
10.5.1 游标概念	(260)
10.5.2 操作游标	(260)
习 题	(267)
习题答案	(268)
第11章 存储过程与触发器	(269)
11.1 存储过程	(269)
11.1.1 存储过程的功能及优势	(269)
11.1.2 存储过程类型	(269)
11.1.3 常用系统存储过程	(270)
11.1.4 设计存储过程	(276)

11.1.5 实现存储过程	(277)
11.2 触发器	(285)
11.2.1 DML 触发器	(285)
11.2.2 DDL 触发器	(295)
习 题	(299)
习题答案	(300)
第12章 数据库安全管理	(301)
12.1 事务	(301)
12.1.1 事务特性	(301)
12.1.2 事务管理	(302)
12.2 SQL Server 的安全机制	(305)
12.2.1 安全机制级别	(305)
12.2.2 主体	(306)
12.2.3 SQL Server 中的身份验证	(306)
12.2.4 数据库用户	(308)
12.2.5 角色	(309)
12.3 SQL Server 的权限管理	(312)
12.3.1 权限类型	(312)
12.3.2 设置权限	(313)
习 题	(315)
习题答案	(316)
第13章 备份与还原数据库	(317)
13.1 备份数据库	(317)
13.1.1 备份与还原	(317)
13.1.2 备份概述	(321)
13.1.3 创建备份	(322)
13.2 还原数据库	(327)
13.2.1 还原数据库方案	(327)
13.2.2 实施还原方案	(328)
13.3 导入导出大容量数据	(334)
13.3.1 导入导出向导	(335)
13.3.2 复制数据库	(337)
13.4 分离和附加数据库	(338)
13.4.1 分离数据库	(339)
13.4.2 附加数据库	(340)
习 题	(340)
习题答案	(342)
附录 数据库设计案例	(343)
参考文献	(355)

第1章 数据库系统概论

本章主要介绍数据库系统的基本概念,包括数据与信息;数据管理技术发展的三个阶段及其特点;数据模型的两个层次;数据库系统的三级模式、两级映像及两级独立性的体系结构;数据库系统的组成。

1.1 信息、数据与数据处理

在科学、技术、经济、文化和军事等各个领域会遇到大量的数据,如何科学地管理数据是一个重要的课题。数据库技术是使用计算机来管理数据的科学技术,经过多年的研究和实践,数据库技术已发展成为一门完整的学科,运用数据库技术科学、有效地管理数据,已经成为计算机应用的重要领域。

1.1.1 信息与数据

计算机的出现,开创了数据处理的新纪元。数据处理的基本要素是数据的组织、存储、检索、维护和加工利用,这些正是数据库系统所要解决的问题。

数据是数据库系统研究和处理的对象。数据与信息是分不开的,它们既有联系又有区别。

1. 信息

随着社会的发展和科学技术的进步,人们对信息这个名词已经不再陌生。对于信息的定义,从不同角度又有着不同的解释。一般认为,信息是人们进行各种活动所需要的知识,是现实世界各种状态的反映。合理利用信息可以增加人们的知识,提高人们对事物的认识能力。在现代社会,不论是生产、科学研究和社会活动,还是个人生活,都离不开信息。

2. 数据

数据是描述信息的符号,也是信息的载体。如数值、文本、图形、图像、声音等类型的数据,用来反映不同类型的信息。利用计算机进行信息处理,就需要把信息转换为计算机能够识别的符号,即用0和1两个编码符号来表示各种信息。

数据和信息既有联系又有区别。如果把客观世界的某种现象或观念所反映的知识用一定的方法描述出来,那么前者是信息,后者是数据。因为信息和数据都是通过现象和概念反映知识的,这是它们的共同点。因此当不需要严格区分时,这两者是一样的。

信息以数据的形式处理,而处理的结果又可能产生新的信息。

1.1.2 数据处理

数据处理是指对各种形式的数据进行收集、存储、加工和传播等一系列活动的总和。其目的是从大量的、原始的数据中抽取、推导出对人们有价值的信息,以作为行动和决策的依据;数据处理从根本上来说是为了借助计算机科学地保存和管理复杂的、大量的数据,以便人们能方便且充分地利用这些宝贵的信息资源。

在数据处理的一系列活动中,数据的收集、组织、存储、传播、检索、分类等活动是基本环节,这些基本环节统称为数据管理或信息管理。在数据处理中,数据的加工、计算、打印报表等操作对不同的业务部门可以有不同的内容。数据库技术属于数据管理技术。数据库技术所研究的问题就是如何科学地组织和存储数据,如何高效地获取和处理数据。

1.2 数据管理技术的发展

数据管理技术的发展与硬件(主要是外存)、软件、计算机应用的范围有着密切的联系。数据管理技术的发展经过三个阶段:人工管理阶段、文件系统阶段和数据库系统阶段。

1.2.1 人工管理阶段

人工管理阶段是从20世纪40年代中期电子计算机问世到50年代中期,这一阶段的计算机系统中,既没有操作系统,也没有管理数据的软件。计算机主要用于科学计算。

在人工管理阶段,数据管理的特点如下。

①数据不保存在机器中。因为计算机主要用于科学计算,一般不需要将数据长期保存。在计算时将数据输入,计算完毕将数据输出。

②没有软件系统对数据进行管理。程序员不仅要规定数据的逻辑结构,而且还要在程序中设计物理结构,包括存储结构、存取方法、输入输出方式等。因此程序中存取数据的部分要随着存储结构的改变而改变,使得数据与程序不具有独立性,这样程序员不仅在数据的物理布置上必须花费许多精力,而且一旦数据在存储结构上有改变,就必须修改程序。

③只有程序的概念没有文件的概念。数据的组织方式必须由程序员自行设计。

④数据是面向应用的。一组数据对应一个程序,即使两个应用程序涉及某些相同的数据,也必须各自定义,所以程序与程序之间有着大量重复的数据,如图1.1所示。



图 1.1 人工管理阶段程序与数据的关系

1.2.2 文件系统阶段

从20世纪50年代中期到60年代中期是文件系统阶段。这一阶段计算机不仅用于科学计算,还大量用于管理。计算机硬件比过去有了较大的发展,外存储器有了磁盘、磁鼓等直接存取的存储设备。在软件方面,操作系统中已经有了专门的数据管理软件,一般称为文件系统。

1. 文件系统阶段数据管理的特点

(1)数据可以长期保存在外存储设备上

由于计算机大量用于数据处理,数据需要长期保留在外存储设备上反复处理,即进

行查询、修改、插入和删除等操作。

(2)数据的逻辑结构与物理结构的区别

由于有了文件系统,文件的逻辑结构与存储结构由系统进行转换,使程序与数据之间有了一定的独立性。这样程序员可以集中精力研究算法,而不必过多地考虑物理细节。

(3)文件组织呈现多样化

由于已有了直接存取存储设备,也就有了索引文件、链接文件和直接存取文件等。

(4)数据不再属于某个特定的程序,可以重复使用

由于文件结构的设计仍然基于特定的用途,程序基于特定的存储结构和存取方法,因此程序与数据结构之间的依赖关系并未根本改变。

在文件系统阶段,由于设备具有独立性,因此改变存储设备,也不必改变应用程序。但还未能彻底体现数据的逻辑结构独立于物理存储结构的要求。当数据的物理结构需要修改时,仍需修改应用程序。

2. 文件系统结构的缺陷

随着数据管理规模的扩大,数据量急剧增加,文件系统结构显露出以下缺陷。

(1)数据冗余度大

相同的数据存在多份的现象,称为数据冗余。文件系统中的文件基本上是对应于某个应用程序的,即数据还是面向应用的。即使不同的应用程序所需要的数据有部分相同时,也必须建立各自的文件,而不能共享相同的数据,因此数据冗余度大,浪费存储空间,并且由于相同数据的重复存储、各自管理,给数据的修改和维护带来了困难,容易造成数据的不一致性。

(2)数据和程序缺乏独立性

文件系统中的文件是专属于某一特定应用服务的,一旦数据的逻辑结构改变,则必须修改应用程序和文件结构的定义。而应用程序的改变,也将影响文件数据结构的变化,即数据和程序缺乏独立性。这个阶段程序与数据的关系如图 1.2 所示。

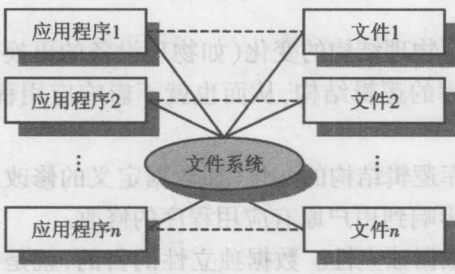


图 1.2 文件系统阶段程序与数据的关系

(3)数据间联系弱

由于文件之间是独立的,要想在文件之间建立联系必须通过程序完成。文件系统是一个无弹性的、无结构的数据集合,不能反映现实世界事物之间的联系。

随着人们对数据处理需求的增加,以及计算机科学的不断发展,对数据进行有效、科学、正确、方便的管理成为人们的迫切需求。针对文件系统的缺陷,人们逐步发展了以统一管理 and 共享数据为主要特征的数据库管理系统。

1.2.3 数据库系统阶段

到20世纪60年代后期,计算机的软硬件得到了进一步的发展,已配备了速度高、容量大的磁盘,各种软件系统进一步完善,需要管理的数据量急剧增加。人们在数据管理方面已积累了丰富经验,数据管理技术研究取得了很大进展,这为数据库系统的研究提供了良好的技术基础。

1968年美国IBM公司成功开发了世界上第一个数据库管理系统IMS(Information Management System);1969年美国CODASYL(Conference On Data System Language)委员会的DBTG(DataBase Task Group)小组公布了它的研究成果DBTG报告;1970年IBM公司的研究员E. F. Codd发表了题为《大型共享数据库数据的关系模型》等一系列关系数据库论文。这三大事件标志着数据处理已进入了数据库技术的新时代。

数据库系统克服了文件系统的缺陷,提供了对数据更高级更有效的管理。

1. 数据库系统阶段的特点

(1) 数据是结构化的

这就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。文件系统中尽管记录内部已经有了某些结构,但记录之间是无联系的,孤立的。因此数据的结构化是数据库的主要特征之一,也是数据库与文件系统的根本区别。

(2) 数据冗余度小,易扩充

由于数据库是从整体观点来看待和描述数据的,数据不再是面向某一应用,而是面向整个系统,这就大大减少了数据的冗余度,既节约存储空间,减少存取时间,又可避免数据之间的不相容性和不一致性。

(3) 具有较高的数据独立性

所谓数据的独立性,就是应用程序不必因为数据的存储结构的变化而修改,即应用程序与数据结构之间不存在依赖关系,这是数据库系统所努力追求的一个目标。数据库的数据独立性分为两级。

①物理独立性。数据库物理结构的变化(如物理设备的更换,物理位置的变化,存取方法的改变等),不影响数据库的逻辑结构,从而也就不影响应用程序,不会导致应用程序的修改。

②逻辑独立性。数据库逻辑结构的变化(如数据定义的修改,新数据类型的增加,数据间联系的变更等等),不会影响到用户原有应用程序的修改。

这两种独立性统称为数据独立性。数据独立性的目的,就是使应用程序尽可能不受数据的影响。这两种数据独立性是靠数据库管理系统来实现的,从而大大减轻了程序员的责任。

(4) 统一的数据控制功能

数据库是不同用户的共享资源。计算机的共享一般是并发的,即许多用户同时使用数据库。因此,系统必须提供以下数据控制功能。

①数据的安全性控制。数据的安全性是指保护数据以防止不合法的使用造成数据的泄密和破解。这就要求采取一定的安全保密措施。例如,系统用检查口令或其他手段来检查用户身份,合法用户才能进入数据库系统。同时,要求提供用户访问级别和数据存取权限,

当用户执行数据库操作时,系统自动检查用户是否有权执行这些操作,检查通过后才执行允许的操作。

②数据的完整性控制。数据的完整性是指数据的正确性、有效性与相容性。系统提供必要的功能,保证数据库中的数据在输入、修改过程中始终符合原来的定义和规定。例如:学生性别只能是男或女,学号是唯一的等。

此外,当计算机软、硬件发生故障而破坏了数据或对数据库数据的操作发生错误时,系统能进行应急处理将数据库恢复到正确状态。

③并发控制。当多个用户的并发进程同时存取、修改数据库时,可能会发生互相干扰而得到错误的结果,导致数据库的完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

(5)数据的最小存取单位是数据项

数据库中数据的存取单位可以是一个数据项或一组数据项,也可以是一条记录或一组记录。这个阶段程序和数据的关系如图 1.3 所示。

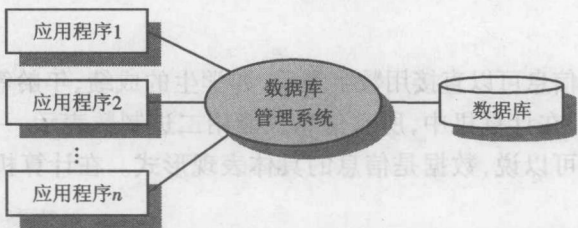


图 1.3 数据库系统阶段程序与数据的关系

数据库是一个通用化的综合性数据集合,它可以供各种用户共享且具有最小的数据冗余度和较高的数据与程序的独立性。由于多种程序并发地使用数据库,为了能及时有效地处理数据,并保证数据库的安全性和完整性,必须有一个软件系统——数据库管理系统(DataBase Management System, DBMS)在建立、运用和维护时对数据库进行统一控制。

1.3 数据模型

1.3.1 数据描述的三个领域

数据库中所存储的数据来源于现实世界的信息流,是用来描述现实世界中一些事物的某些方面的特征及其相互联系的。在处理这些信息之前,必须先分析它,选择一种方法描述这些待处理对象,并将这种描述转换成计算机所能接收的数据形式。在数据处理中,数据描述涉及不同的范畴。从事物的特性到计算机中的数据表示,都涵盖了现实世界、信息世界和机器世界三个领域。

1. 现实世界

现实世界是指存在于人脑之外的客观世界,泛指客观存在的事物及其相互间的联系。客观事物是一个具体的事物,如一个学生、一台计算机、一本书等,也可以是一个抽象的事物,如一次比赛、一次借书等。

每个客观事物都有自己的特征,用来区别于其他客观事物,如学生用姓名、性别、年龄、

身高、体重等许多特征来标识自己。把具有相同特征的客观事物称为同类客观事物,所有同类事物的集合称为总体。例如所有的“学生”、所有的“课程”都是一个总体。

所有这些客观事物是信息的源泉,也是数据库技术接触到的最原始的数据,数据库设计是对这些原始数据进行综合处理,抽取出数据库技术所需要的数据。

2. 信息世界

现实世界中的事物反映到人们的头脑里,经过认识、选择、命名、分类等综合分析而形成印象和概念,产生认识,这就是信息,即进入了信息世界。在信息世界中,每一个被认识的客观事物称为实体,是具体事物在人们头脑中产生的概念,是信息世界的基本单位。另外,客观事物的特征称为属性,属性反映了实体的某一特征。一个实体是由它所有的属性表示的。例如,一本书是一个实体,可以由书号、书名、作者、出版社、单价 5 个属性表示。

在信息世界里,主要研究的不是个别的实体,而是它们的共性。把具有相同属性的实体称为同类实体,同类实体的集合为实体集。例如,所有的男学生组成了男学生实体集。能唯一标识每个实体的属性或属性集称为实体标识符。例如,书的书号可以作为书的实体标识符。

3. 机器世界

信息世界中,有些信息可以直接用数字表示,如学生的成绩、年龄等;有些是由符号、文字或其他形式表示的。在计算机中,所有信息只能用二进制数表示,一切信息进入计算机时,必须是数据化的。可以说,数据是信息的具体表现形式。在计算机世界中涉及以下术语。

- ①数据项。数据项是属性的数据表示,它是可以命名的最小信息单位,又称数据元素或字段。如职工的职工号、姓名等。
- ②记录。记录是实体的数据表示,由若干个数据项组成,如一个职工就是一条记录,它由职工号、姓名、性别、职称等数据项组成。
- ③文件。文件是同类记录的集合,如所有职工的登记表组成一个文件。
- ④关键字。能唯一标识文件中每条记录的数据项或数据项的集合,称为记录的关键字(或“键”)。如职工的职工号可以作为职工记录的关键字。

现实世界、信息世界、机器世界是由客观到认识、由认识到使用管理的不同领域,而且后一领域是前一领域的抽象描述。

三个领域之间的术语对应关系如图 1.4 所示。

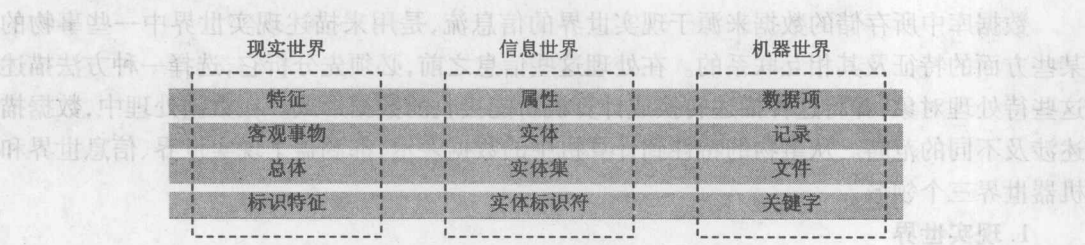


图 1.4 三个领域之间的术语对应关系

1.3.2 数据模型

1. 数据模型的基本概念

模型是对现实世界特征的模拟和抽象。数据模型是对现实世界数据特征的抽象。在数据库中用数据模型对现实世界进行抽象,数据模型是数据库系统中用于提供信息表示和操作手段的形式构架。

在数据库系统中针对不同的使用对象和应用目的,采用不同的数据模型。不同的数据模型提供了不同的模型化数据和信息工具。根据模型应用的不同目的,可以将模型分为概念数据模型(也称信息模型)和结构数据模型两类。

(1) 概念数据模型

用于信息世界的建模,它是从数据的语义视角来抽取模型并按用户的观点对数据和信息建模,是现实世界到信息世界的第一层抽象,是用户和数据库设计人员之间进行交流的语言。这类模型主要用于数据库的设计阶段,常用的概念数据模型是实体联系模型。

(2) 结构数据模型

用于机器世界,它是从数据的组织层次来描述数据并按计算机系统的观点对数据建模。结构数据模型可以用来定义、操纵数据库中的数据,是信息世界到机器世界的第二层抽象,主要用于数据库管理系统的实现,包括层次模型、网状模型、关系模型和面向对象模型等。

2. 数据描述的概念

(1) 数据的静态描述

它包括数据的基本结构、数据间的联系和数据中的约束。例如学生的基本信息包括学号、姓名、性别、年龄、专业,这些都是数据静态结构中的基本信息。学生进行选课时使用的学号必须是学生基本信息中的学号,这就是数据间的联系。学生的性别只能取“男”或“女”,这就是数据中的约束。

(2) 数据的动态描述

它指定在数据上的操作,例如对学生基本信息进行修改、查询、增加和删除记录等操作。

3. 数据模型的三要素

这里所说的数据模型指结构数据模型。数据模型是严格定义的概念集合,这些概念精确地描述系统的静态特性、动态特性和完整性约束条件。因此,数据模型通常由数据结构、数据操作和数据约束条件三部分组成。

(1) 数据结构

数据结构是对系统静态特性的描述,是所研究的对象类型的集合,这些对象是数据库的组成成分。一般可分为两类:一类是与数据类型、内容、性质有关的对象,在关系模型中对应的是域、属性、关系等;另一类是与数据之间的联系有关的对象。

数据模型中的数据结构用以构造数据库的基本数据结构类型,并规定如何把基本数据项组成大的数据单位,以及如何表达数据项之间的联系。在数据库系统中通常按照数据结构的类型来命名数据模型,如层次结构、网状结构和关系结构的模型分别命名为层次模型、网状模型和关系模型。

(2) 数据操作

数据操作是指对数据库中各种对象(型)的实例(值)允许执行的操作的集合。数据操